arXiv:2303.01501v1 [stat.CO] 2 Mar 2023

使用 Delaunay-Rips复形的持久同调稳定性与机器学习应用

Amish Mishra 和 Francis C. Motta

摘要

本文定义、实现并研究了一种用于计算欧几里得点云数据持久同调的单纯复形构造方法，称为Delaunay-Rips复形（DR）。为单形赋予 Vietoris-Rips权重后，DR 仅考虑出现在点云 Delaunay三角剖分中的单形，从而加快持久性计算。我们介绍并比较了 DR 的 Python 实现与其他用于生成持久图的单纯复形构造方法。通过对点云数据施加充分条件，我们从理论上证明了使用 DR 生成的持久图具有稳定性。当点云的 Delaunay三角剖分因点的扰动而发生变化时，我们证明 DR 生成的持久图会表现出不稳定性。由于无法保证现实世界数据满足我们的稳定性条件，我们将 DR 用于持久同调时的实际稳健性与机器学习应用中的其他单纯复形进行比较。实验结果表明，在机器学习拓扑数据分析流程中使用 DR，效果与使用其他单纯复形构造方法相当。

目录

1 引言 **1**

2 背景 **3**
2.1 单纯同调.3 2.2 持久同调.5 2.3 Vietoris-Rips 和 Alpha复形.6

3 Delaunay- Rips 复形与稳定性 7
3.1 定义与构造.7 3.2 实现与运行时间分析.9 3.3 Delaunay-Rips复形的稳定性性质.11 3.4 持久图不稳定性.13

4 机器学习 模型性能：使用Rips、Alpha和Delaunay-Rips过滤 15
4.1 合成形状数据分类.16 4.2 睡眠状态分类.17

5 结论 **20**

完整参考文献 **21**

1 引言

随着实验科学和工业应用中收集的数据量及其复杂性不断增长，科学家和工程师往往会采用一些方法，将数据转换为更紧凑、易于管理的表示形式，同时保留关键特征，以便开展有意义的数据分析和可视化，并支持开发高性能预测模型。应用计算拓扑学分支——拓扑数据分析（TDA）中的工具，提供了多种解决方案，可应对复杂或高维数据处理与分析中的部分挑战。TDA 利用代数拓扑的研究成果，测量并量化数据中定性、基于形状的特征，并持续

吸引着计算、数学和实验学科领域研究人员的兴趣，包括金融网络 [Lei+08]、图像信号 [CBK09], 癌症检测 [Qai+16]、法医学 [AJ17], 以及材料科学 [Kra+14]。

持久同调（PH）是 TDA 中的旗舰方法，提供了一种稳健的方法，可将数据集的多维几何/拓扑特征编码为一种称为持久图（PD）的紧凑表示。PH 通常根据点云数据计算，即度量空间中的有限点集。首先，为点云关联一个嵌套的拓扑空间族——也就是由不同维度的单形（例如点、边、三角形、四面体等）组成的单纯复形集合——该集合由一个（实数）尺度参数确定。这种单纯复形的过滤旨在捕捉点云在不同尺度下的几何/拓扑结构。根据该过滤，可得到一个持久图，由 $\mathbb{R}^2$ 中对角线以上的一组有序对组成。每个（出生、消亡）对表示一个固定维度的拓扑特征（连通分支、孔洞、空腔等），该特征在出生尺度出现，并在消亡尺度消失。

构造过滤取决于两个因素：哪些单纯形应纳入复形族，以及确定每个单纯形应在哪个尺度出现。这两个因素都会增加在点云数据上计算持久同调的总体计算负担，也可能成为将持久同调用于大型或高维数据集的障碍。例如，一种颇具吸引力且常用的构造方法称为 Rips 过滤，它只需知道点与点之间的成对不相似度。因此，确定所有单纯形出现的尺度相当直接。然而，Rips 复形的大小会随点的数量呈指数级增长。Rips 过滤的一种替代方法是 Alpha 复形过滤 [Ede93], ，它将复形限制为点云的 Delaunay三角剖分中的单纯形。不过，该方法根据与每个数据点相关联的沃罗诺伊单元之间的邻接关系来为单纯形指定尺度，而这可能需要大量计算。最近还提出了其他一些解决方案 [She12; GO07; SC04] ，以缩小输入持久同调算法的复形规模。持久同调算法已在多种语言（C++、Python、Julia 等）中实现，并经过优化以进一步提升计算效率 [Bau21; TSB18; ˇCuf20]。

在本研究中，我们受到启发，将 Rips 过滤与 Alpha 过滤的计算效率结合起来，为欧几里得点云数据定义一种新的过滤。我们将这一构造称为 Delaunay-Rips（DR）复形，它可视为惰性见证复形 [SC04] 的一个特例。我们没有使用弱见证复形，而是采用作为 Delaunay三角剖分[Sil08] 的强见证复形。我们还证明，一般而言，这种过滤将继承先前确立的稳定性结果；这些结果保证，粗略地说，底层数据发生某些小变化时，持久图 [Cha+09；CEH05] 也会发生小变化。尽管如此，我们进一步证明，对于某些点的构型，DR 过滤会出现不稳定性：底层数据的任意小扰动都可能导致持久图发生非无穷小变化。这种不稳定性源于点处于退化构型时，Delaunay三角剖分本身的不稳定性。

PH 技术的许多根本吸引力源于大量稳定性结果，这些结果表明，将数据映射为图和/或将图映射为向量化拓扑表示的变换是利普希茨连续变换 [ST20]。然而，在实践中，由 PDs 得到并用于数据分析和统计模型开发的拓扑表示往往并不稳定，尽管基于这些表示的模型据称取得了成功。因此，我们进一步希望探究：实践中，DR 过滤中观察到的不稳定性究竟有多大影响？我们通过在合成数据集和真实数据集以及机器学习（ML）建模任务上，对 Alpha、Rips 和 DR 过滤的性能与成本进行系统比较，以实证检验这一问题。我们发现，在这些情况下，不稳定性的影响可以忽略不计，而 DR 所具有的实际计算优势可能十分显著。因此，本文的贡献如下：

- 我们定义了 DR 复形，它以德劳内三角剖分为骨架，对数据上的 Vietoris-Rips 单形进行了细化。
- 我们提供了一种 DR 过滤算法及其 Python 实现。
- 我们记录了使用 DR 构造持久图的若干实测运行时间比较，并将其与 Rips [TSB18] 和 Alpha [TS] 过滤的常用实现进行了比较。

- 我们给出了一份直接的证明，说明在一般位置点云的某个邻域内，Delaunay三角剖分具有稳定性；这也确立了 DR 过滤在数据的适当选定邻域内的稳定性。

- 我们给出了一个罕见的证明，揭示了用于计算持久同调的某种过滤存在不稳定性。
- 我们考察了这种不稳定性对合成数据和真实世界分类任务的实际影响，发现它对模型性能的影响有限。

本文结构如下：第 2 节简要回顾必要的数学预备知识。第 3 节正式介绍德劳内-Rips 过滤，提供了我们用于计算点云数据的德劳内-Rips 过滤的实现伪代码，并比较了该算法与 Rips 和 Alpha 构造的实现，在点云规模和维度不断增大时的实测运行时间。第 3.3 节讨论德劳内-Rips 过滤的稳定性性质，并通过手工计算一族持久图，展示在底层点云受到扰动时，点云到持久图的变换如何出现不连续性。第 4 节报告了若干系统性比较的结果：在合成数据和实验数据上，比较基于 Alpha、Rips 和德劳内-Rips 过滤得到的拓扑特征训练的机器学习模型的性能，其中包括使用持久图的持久图像（PI）向量化表示训练的随机森林分类器 [Ada+17] ，以及使用从心电图时间序列数据得出的持久性统计特征向量训练的支持向量机 [Chu+21]。

2 背景

要从点云中提取拓扑特征，通常首先将这些点视为所谓单纯复形的顶点。单纯复形由顶点、边（由两个顶点组成）、三角形（由三个顶点组成）、四面体（由四个顶点组成）以及更高维的类似结构（由 $n > 4$ 个顶点组成）构成，从而构造出一个具有明确定义的“形状”概念的对象。我们关注的是计算代数对象，即持久同调群，它们能够刻画跨尺度的拓扑（同调）不变量（例如连通分支、洞、空腔等）。这里我们简要回顾这些概念。关于本节数学基础的深入讲解，请参阅 [EH10；Hat02]。若要了解更多计算方面的内容，包括诸多实际考量和示例，我们推荐 [Ott+17]。

2.1 单纯同调

定义 2.1。设 K_0 是一个有限集合， $\mathcal{P}(K_0)$ 是 K_0 的幂集（即所有子集构成的集合）。基于 K_0 构建的抽象单纯复形是由 K_0 的非空子集构成的一个集合 $K \subset \mathcal{P}(K_0)$，并满足以下性质：对于所有 $v \in K_0$，都有 $\{v\} \in K$；且若 $\sigma \in K$，则对于所有 $\tau \subseteq \sigma$，都有 $\tau \in K$。

本文关注由点云构建的单纯复形，其中点与复形中的单元素集合相对应。一般而言，单纯复形的元素 K 称为单形，而足够小的子集也可能用其他名称指称。例如，我们将单元素集合称为 K 的顶点，将大小为二的集合称为边，等等。如果一个单形的大小为 $p+1$，我们就说它的维度为 p，或者称它为 p-单形；因此，顶点是维度为 0 的单形，边是维度为 1 的单形，三角形是 2-单形，等等。若一个 p-单形包含 0-单形 v_i、$i = 0$、$\ldots p$，我们用 $[v_0 v_1 \ldots v_p]$ 表示它。K_p 表示所有 p-单形构成的集合，而 K 的 k-骨架是对所有 $p \in \{0, 1, \ldots, k\}$ 取集合 K_p 的并集。如果 τ 和 σ 是满足 $\tau \subset \sigma$ 的单形，那么我们称 τ 为 σ 的一个面。如果 τ 是 σ 的一个余维为 q 的面，那么 τ 和 σ 的维度相差 q。K 的维度定义为其所有单形维度的最大值。

定义 2.2. 设 K 是一个单纯复形，且 $p \geq 0$。K 的 p-链是 K 中 p-单形的形式和，写作

$$c = \sum a_i \sigma_i$$

其中 $a_i \in \mathbb{F}$ 是一个域， σ_i 是 p-单形。于是， p-链构成了域 $\mathbb{F}$ 上由 p-链组成的向量空间，我们将其记为 $C_p(K)$。

$C_p(K)$ 和 $C_{p-1}(K)$ 通过一个称为边界映射的线性映射自然关联，该映射将每个 p-链映射为其边界（$p-1$）-链。

定义 2.3. 设 K 是一个 n 维单纯复形，且 $\sigma = [u_0 u_1 \dots u_p] \in C_p(K)$。$p$ 边界映射 ∂_p 定义在 p 单纯形上，定义为

$$\partial_p \sigma = \sum_{j=0}^{p} (-1)^j [u_0 \dots \hat{u}_j \dots u_p]$$

其中帽子表示省略 u_j。通过线性性扩展，$\partial_p : C_p(K) \to C_{p-1}(K)$ 进一步定义在任意 p 链 $c = \sum_i a_i \sigma_i$ 上，定义为

$$\partial_p c = \sum_i a_i \partial_p \sigma_i.$$

直观地说，单纯形的余维 1 面所构成的链刻画了该单纯形的边界。来看一些例子。三角形 $[abc]$ 的边界是链 $[bc] - [ac] + [ab]$，由构成其边界的三条边组成。扩展到链上，与顶点 b 相连的两条边的边界（即 $[ab] + [bc]$）为 $[b] - [a] + [c] - [b] = [c] - [a]$，这体现了这个一维对象的边界仅由其两个端点构成。通过边界映射连接链群，就构成了一个链复形

$$\mathbf{0} \to C_n(K) \xrightarrow{\partial n} \dots C_{p+1}(K) \xrightarrow{\partial p+1} C_p(K) \xrightarrow{\partial p} C_{p-1} \xrightarrow{\partial p-1} \dots C_0(K) \to \mathbf{0},$$

其中，第一个映射是平凡线性映射，它将平凡向量空间中的零向量 0 映射到 $C_n(K)$ 中的零向量。

非正式地说，单纯复形中的一个 p 维孔洞，将由一个可能是（但并非）某个 p 链边界的 $(p-1)$-链表示。例如，正如我们所见，无边界的 1-链 $[bc]+[ac]+[ab]$是一个 2-链的边界（即 2-单纯形 $[abc]$）。然而，如果 $[abc]$ 不在该单纯复形中，我们就可以说，这个复形包含一个由缺失三角形的边围成的孔洞。为了精确定义这些概念，我们引入 $C_p(K)$ 的两个子空间，它们分别编码所有没有边界的 p-链，以及实际上是 $p+1$ 链边界的那些链。

定义 2.4。$\gamma \in C_p(K)$ 中的一个 p-圈，是一个没有边界的 p-链，即 $\partial_p \gamma = 0$。单纯复形 K 中所有 p-圈的集合记作 $Z_p(K)$，并构成 $C_p(k)$ 的一个子空间，因为 $Z_p(K) = \ker \partial_p$.

定义 2.5。p-边界 $\beta \in C_p(K)$ 是一个 p-链，它是某个 $(p+1)$-链的边界，即，对于某个 $\sigma \in C_{p+1}(K)$，有 $\partial_{p+1}\sigma = \beta$。单纯复形 K 中所有 p-边界的集合记为 $B_p(K)$；由于 $B_p = \text{im } \partial_{p+1}$，它也是 $C_p(K)$ 的一个子空间。

定义了 p-圈和 p-边界后，我们可以正式定义一种对象，用来表示单纯复形中的孔洞：那些不是边界的圈。

定义 2.6。第 p 阶同调群为 $H_p = Z_p(K)/B_p(K) = \ker \partial_p / \text{im } \partial_{p+1}$.。第 p 个贝蒂数是该商向量空间的维度，即 $\beta_p = \text{rank } H_p$。

H_p 因此是 p-圈的等价类所构成的向量空间，其中两个 p-圈等价，当且仅当它们相差一个 p–边界。H_p 是良定义的，这依赖于一个基本结果：边界链的边界必为空，因此 B_p 实际上是 Z_p的一个子空间。

引理 2.7。（同调基本引理 *[EH10]*）链复形中任意两个连续边界映射的复合都是平凡的。也就是说，

$$\partial_p \partial_{p+1}(\sigma) = 0$$

对于每个整数 p 以及每个 $(p+1)$-链 σ。

如前所述，第 p 个贝蒂数对应于相应单纯复形中 p 维“孔洞”的数量。事实上，0 阶贝蒂数统计连通分支的数量，1 阶贝蒂数统计（相互独立的）环的数量，2 阶贝蒂数统计（相互独立的）空腔数量，依此类推。

2.2 持久同调

尽管单纯同调足以捕捉固定单纯复形的一些内在形状特征，但对于点云数据，自然会引出这样一个问题：如何根据这些点构造单纯复形，从而有意义地表示点云中的潜在结构。事实上，对于基于数据构造的某个复形，可能会存在感兴趣的拓扑特征（例如在某个较精细的尺度下）；而对于另一个复形，也可能会有另一组值得关注的拓扑特征（在更大的尺度下）。解决这一问题的一种有前景的方法是持久同调。它是同调的一种扩展，考察的是单纯复形的参数化族，而非单个复形。

定义 2.8. 对每个 $s \in \mathbb{R}$，令 K_s 为一个单纯复形，并且对所有 $s \leq t$，满足 $K_s \subset K_t$。我们将这样一组复形称为过滤。当每个 $K_s \subseteq K$ 对某个固定的有限单纯复形 K 成立，且对某个 $t \in \mathbb{R}$， $K_t = K$ 成立时，我们将这个参数化集合称为 K 的过滤。

注意，有限单纯复形 K 上的过滤必然只包含有限个不同的子复形，我们可以重新标记

$$\emptyset := K_0 \subseteq K_1 \subseteq K_2 \subseteq \cdots \subseteq K_n = K.$$

这里我们将 K_i 重新标记为：对于子复形发生变化时对应的尺度 $s_i \leq s_{i+1}$， $= K_{s_i}$。每个单形 $\sigma \in K$ 也可以被赋予其在过滤中出现的最小尺度。换言之，可以定义 $f: K \to \mathbb{R}$，使得 $f(\sigma) = t$ 当 $\sigma \in K_t$，并且对于任意 $s < t$， $\sigma \notin K_s$。因此，过滤在 K 的单形上诱导出一个所谓的单调函数，其中若 σ 是 τ 的面，则 $f(\sigma) \leq f(\tau)$。反过来，任何满足如下性质的单调函数 $f: K \to \mathbb{R}$，若 $f(\sigma) \leq f(\tau)$ 当 $\sigma \subseteq \tau$，则可通过取 $K_s = f^{-1}(\infty, s]$. 在 K 上定义一个过滤

定义 2.9. 设 $\emptyset \subseteq K_1 \cdots \subseteq K_n = K$ 是 K的一个过滤。该过滤的第 p阶持久同调群形式化定义如下：

$$H_p^{i,j} = Z_p(K_i) / \left(B_p(K_j) \cap Z_p(K_i)\right).$$

对于 $0 \leq i \leq j \leq n$。相应的第 p阶持久贝蒂数是这些群的秩，

$$\beta_p^{i,j} = rank\ H_p^{i,j}.$$

$H^{i,}{}_p{}^j$中的一个元素对应于过滤中一个从 K_i 持续到 K_j 的圈（即在 K_i 中尚未成为边界的圈）；过滤的所有第 p阶持久同调群的内容可概括为一个维度为p 的持久图，该图追踪同调特征在过滤中首次出现和随后消失时所对应的指标对。通常将一个特征在复形 K_i 中“诞生”、在复形 K_j 中“消亡”，意指它在 K_i 的尺度上成为一个圈（但不是边界），并且在 K_j中最早成为边界。

定义 2.10. 设 $\mu^{i,}{}_p{}^j$表示在复形 K_i 中诞生、进入复形 K_j时消亡的 p维类的数量。那么

$$\mu_p^{i,j} = (\beta_p^{i,j-1} - \beta_p^{i,j}) - (\beta_p^{i-1,j-1} - \beta_p^{i-1,j}).$$

p-持久图由过滤 $f: K \to \mathbb{R}$给出，记作 $\mathrm{Dgm}_p(f)$，它是扩展实平面 $\bar{\mathbb{R}}^2$ 中的一组多重对 (s_i, s_j)，其重数为 $\mu^{i,}{}_p{}^j$。每一对 (s_i, s_j) 表示一个非平凡的持久同调类：它在复形 $K_i = K_{s_i}$ 中诞生，并在进入复形 $K_j = K_{s_j}$ 时消亡，因为它与一个在 K_i之前诞生的同调类合并。

现在，我们可以定义一种度量，用来计算两个 p-持久图之间的距离。

定义 2.11。设 X和 Y是两个 p-持久图。定义 $||x-y||_\infty := \max\{|x_1 - y_1|, |x_2 - y_2|\}$，其中 $x = (x_1, x_2) \in X$， $y = (y_1, y_2) \in Y$。我们将这两个图之间的瓶颈距离定义为

$$W_\infty(X, Y) = \inf_{\eta: X \to Y} \sup_{x \in X} ||x - \eta(x)||_\infty$$

其中，下确界取遍所有双射；如果 X 和 Y 的基数不同，则 η 可以映射到对角线。

基于固定复形 K 上的过滤构建的持久图具有一项重要性质：对于确定 K上过滤的单调函数的变化，它们具有稳定性：

定理 2.12（[EH10]）。设 K 为一个有限单纯复形，且 f、g： $K \to \mathbb{R}$ 为两个单调函数。对于每个维度 p，图 $\mathrm{Dgm}_p(f)$ 与 $\mathrm{Dgm}_p(g)$ 之间的瓶颈距离满足

$$W_\infty\left(Dgm_p(f), Dgm_p(g)\right) \leq ||f-g||_\infty,$$

其中 $||f-g||_\infty = \max_{\sigma\in K} |f(\sigma) - g(\sigma)|$.

我们还将使用度量空间中子集之间的距离概念，来控制存在于 $\mathbb{R}^D$ 中的点云之间的距离。

定义 2.13。设 X 和 Y 是度量空间（M，d）中的两个非空子集。我们将它们的豪斯多夫距离 $d_H(X, Y)$ 定义为

$$d_H(X,Y) := \inf\{\varepsilon > 0 \ ; \ X \subseteq Y_\varepsilon \ and \ Y \subseteq X_\varepsilon\},$$

其中

$$X_\varepsilon := \bigcup_{x\in X} \{m \in M \ ; \ d(x,m) \leq \varepsilon\}.$$

对于有限子集 X、$Y \subset$（M、d）（即有限点云），豪斯多夫距离等于一个集合中的点到另一个集合中最近点的最大距离。我们将在引理 3.3 中证明一个相关结论。

本节定义的单纯复形上的过滤及其持久同调群的形式化概念，提供了一种从点云数据中提取多尺度结构的方法，从而在一定程度上缓解了关于哪种复形最能捕捉点云结构的疑虑。不过，从点云构造过滤的方法有很多种，其优缺点取决于数据的性质。

2.3 Vietoris-Rips 和 Alpha 复形

在有限点云 $X \subset (M,d)$ 上构造过滤，最简单且最常用的方法之一，是将点视为顶点，并在由度量空间中的直径确定的尺度下添加单形。

定义 2.14. 设 $X \subset \mathbb{R}^D$ 是一个点云，且 $\varepsilon \geq 0$。尺度为 ε 时的 Vietoris-Rips 复形定义为

$$VR_\varepsilon(X) = \{\sigma \subseteq X \mid d(x,x') \leq 2\varepsilon, \ \forall x, x' \in \sigma\}.$$

换言之，对于给定的尺度 $\varepsilon \geq 0$，若对于 $x, x' \in X$， $d(x,x') \leq 2\varepsilon$，则将 p-单纯形 $\sigma = [x_0 x_1 \dots x_k]$ 添加到复形中，添加尺度取其任一边所对应尺度中的最大值。

尽管在任意尺度下计算 Rips 复形的算法很容易实现，但对包含大量点的点云 X 构造该复形会带来计算难题：最终（在点云直径的一半及更大的尺度下），Rips 复形将等于 X 的幂集，因此会包含 $2^{|X|}$ 个单形。此外，Rips 复形最终会包含所有维度的单形（最高可达点云大小减 1），因此即使超出点云的维度，也会包含同调信息（假设数据位于有限维向量空间中）。在实践中，可以通过限制过滤中任意复形所包含单形的最大维数来缓解这一问题。此外，实践中常见的情况是，随着尺度增大，复形中可能会出现不影响其同调的额外单形 [Mat08；AAM17]。对于大型点云，避免纳入“多余”单形的过滤方法可能更可取 [She12；GO07；SC04]。在定义此类方法的示例之前，为便于理解本文余下内容，先定义欧几里得点云何时处于“一般位置”。

定义 2.15. 如果一个 d 维欧几里得空间中的点集没有 $d+2$ 个点位于同一个 $(d-1)$ 维球面上，则称该点集处于一般位置。

例如，这意味着在 $\mathbb{R}^2$ 中，若一个点集处于一般位置，则其中任意 4 个点都不能共圆。此条件确保每个由 $d+1$ 个点构成的子集都位于唯一的 d 维球面上，从而保证存在唯一的 Delaunay三角剖分，定义如下。

设 $X \subset \mathbb{R}^D$ 是一个有限点云。令 $x \in X$ 并定义

$$V_x := \{p \in \mathbb{R}^D \mid d(p,x) \leq d(p,x') \ \forall x' \in X\}.$$

每个 V_x 都称为 X 的一个沃罗诺伊单元，它包含所有满足以下条件的点：这些点到 X中任意其他点的距离都不小于到 x 的距离。注意，$\{V_x\}_{x\in X}$ 构成了 $\mathbb{R}^D$ 的一个覆盖。这个覆盖称为相对于 X 的 $\mathbb{R}^D$ 的沃罗诺伊分解。为了从这个覆盖构造 Delaunay 三角剖分，我们定义

$$\mathrm{Del}(X) := \{\sigma \subset X \mid \bigcap_{x\in\sigma} V_x \neq \emptyset\}.$$

已知 Del(X) 本身是一个单纯复形 [EH10]。Del(X) 中的一个 n-单纯形 $\sigma \in$ 称为德劳内单纯形。在第 3.1 节定义 Delaunay-Rips 复形时，我们将使用 Del(X) 作为底层结构。

首先，我们回顾另一种常用的过滤构造，称为 Alpha过滤，它在点云 $X \subset \mathbb{R}^D$上得到了深入研究。我们回顾 [EH10] 中 III.4 给出的定义。

定义 2.16。设 $\varepsilon \geq 0$，并令 $S_x(\varepsilon) := V_x \cap B_x(\varepsilon)$，其中 $B_x(\varepsilon)$ 是一个以 $x \in X$为中心、半径为 ε的 d维球。尺度 $\varepsilon \geq 0$ 下的 Alpha 复形为

$$Alpha_\varepsilon(X) = \{\sigma \subseteq X \mid \bigcap_{x\in\sigma} S_x(\varepsilon) \neq \emptyset\}.$$

注意，由于 $S_x(\varepsilon) \subseteq V_x$，$Alpha_\epsilon$ 复形的 1-单形集合构成 Delaunay三角剖分的 1-骨架的子复形。

根据构造，对于所有 $\varepsilon \in \mathbb{R}$，$VR_\varepsilon(X)$ 和 $Alpha_\varepsilon(X)$ 都是单纯复形；并且如果 $s \leq t$，则 $Alpha_s \subseteq Alpha_t$ 和 $VR_s \subseteq VR_t$ 都成立。因此，每种过滤构造都会为基于点云 X 构建的单纯复形中的一组单形确定一个排序。对于 Vietoris-Rips，每个单纯形的尺度由定义该单纯形的两个最远顶点之间的距离决定。然而，Vietoris-Rips 还会为顶点集合的每个子集赋予非零权重，而这会导致计算时间随顶点数量呈指数级增长。另一方面，Alpha 不会计算顶点集合每个子集的尺度。相反，赋予单形的尺度由沃罗诺伊单元上受限的 ε-球何时相交来确定。我们在下一节构造 Delaunay-Rips复形时，正是要避免这项额外计算。

3 德劳内–黎普斯复形与稳定性

3.1 定义与构造

结合 Alpha 和 Rips 构造，可以提供一种在点云 $X \subset \mathbb{R}^d$上构造一族复形的替代方法。这个想法类似于[BE16] 中定义的 Delaunay-˘Cech 复形的构造，也可以视为惰性弱见证复形 [SC04]的一种特殊情形。Delaunay-Rips复形利用了 Vietoris-Rips 复形在概念上的简洁性，同时减少了高维单形以及可能多余的单形数量。其思路是在 X 上构造 Vietoris-Rips 复形，但只有当边出现在点云的德劳内 1-骨架中时，才将其加入。随后按照传统的 Vietoris-Rips 方式添加更高维的 p-单形，即当且仅当它们的 1-骨架出现时才添加。

定义 3.1。给定尺度 $\varepsilon \geq 0$ 时，Delaunay-Rips（DR）复形定义为：

$$DR_\varepsilon(X) = \{\sigma \subseteq Del(X) \mid d(x,x') \leq 2\varepsilon,\ \forall x,x' \in \sigma\}.$$

与此前的 Vietoris-Rips 和 Alpha 过滤一样，Delaunay-Rips 过滤可以视为一种在点云上构建复形并为复形中的单形赋予尺度、从而建立单调顺序的方法，该复形构建于 X之上。

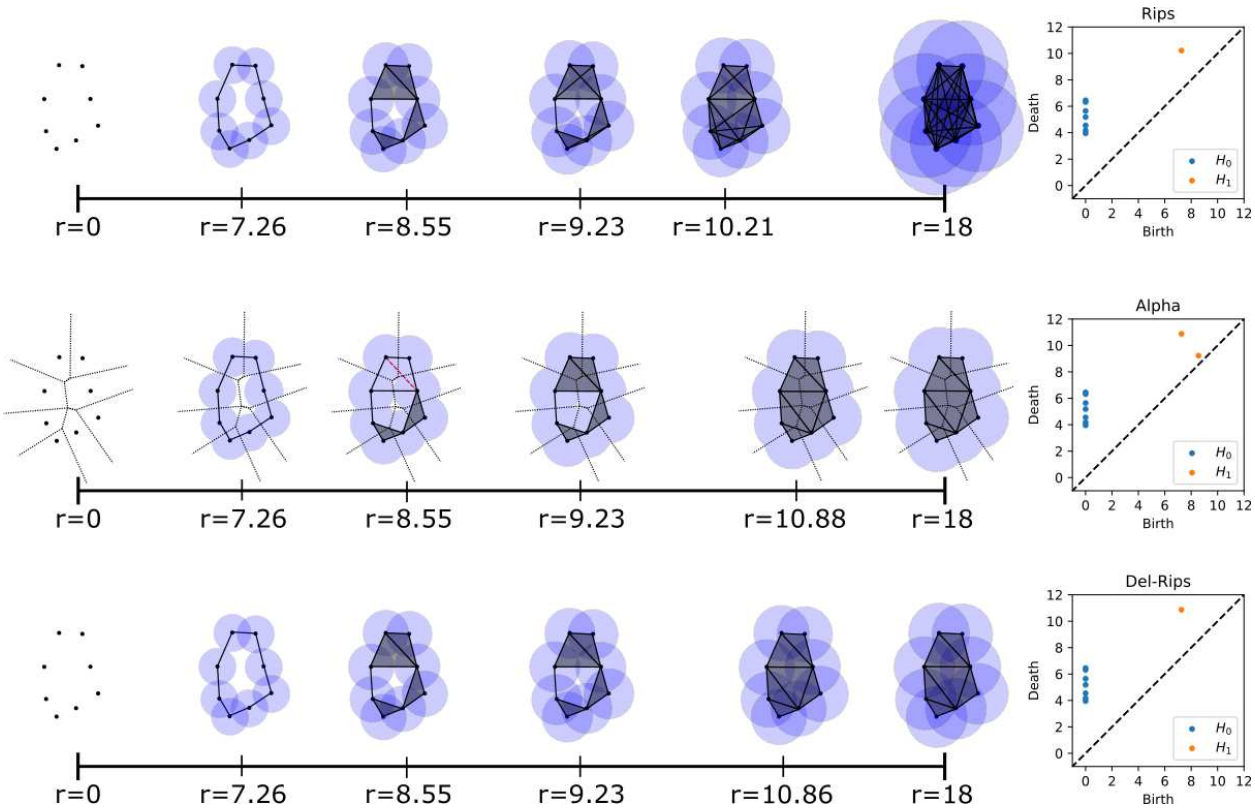


图 1：左侧从上到下依次为点云上的 Rips、Alpha 和 Delaunay-Rips 过滤，该点云包含 $\mathbb{R}^2$ 个点（黑点）。右侧为与各过滤相对应的 H_0 和 H_1 持久图。在每种过滤中，阴影圆在每个尺度值处的半径均为 r；边和三角形则按照相应过滤的定义，在指定尺度下加入。Alpha 过滤中的灰色虚线表示沃罗诺伊单元的边界。在持久图中，蓝色点表示 H_0 类（连通分支），橙色点表示 H_1 类（环）。

图 1 展示并比较了基于 $\mathbb{R}^2$中的示意点云构建的 Rips、Alpha 和 Delaunay-Rips 过滤示例，以及与之对应的 H_0 和 H_1 持久图。Rips 过滤产生了一个在尺度值 7.26 处诞生、并持续到尺度值 10.21 才消亡的同调类。这由持久图中唯一橙色点的诞生和消亡坐标表示。请注意，尽管在尺度值 8.55 时，重叠圆构成的拓扑曲面上有两个环，但 Rips 并未捕捉到上部的第二个环，因为三条边界边一出现，就会加入一个三角形。与此相反，Alpha 过滤在尺度值 7.26 处产生一个同调类，该类持续到尺度值 10.88；此外，还产生另一个 H_1 同调类，它在 8.55 处诞生，并在尺度值 9.23 处消亡。请回想，特定尺度值下的 Alpha 复形要求将$\mathbb{R}^D$ 中一个 d-单形的每个 0-单形为中心的球限制在相应 0-单形的沃罗诺伊单元内。因此，只有当所有受限球都相交时，才会将 d-单形加入复形。在我们的示例中，红色虚线边——它在尺度值 8.55 时与其生成的高维三角形一同出现在 Rips 和 DR 过滤中——由于上述 Alpha 复形的性质，要到尺度值 9.23 才会出现在 Alpha 过滤中。具体来说，这条边必须等到球的半径等于包含构成该单形的三个点的圆的半径时才能出现，而这条边正是该单形的一部分。

DR 过滤产生了一个从尺度值 7.26 持续到 10.86 的 H_1 同调类。在 DR 构造中，一旦复形中出现构成三角形的 3 条边，我们就会添加该三角形：图中用阴影三角形对此进行了示意。请注意，在 $r = 18$ 处，尽管许多圆盘相互重叠，我们只添加了出现在点云的 Delaunay 三角剖分中的单形。最终得到的复形在组合结构上与最终的 Alpha 复形完全相同。这里一个值得注意的对比是，Del-Rips 持久图显示了一个 H_1

类，其出生尺度值与 Rips 过滤中的相同，但持续时间更长（这是由于单形在组合结构上有所减少）。另一个值得注意的对比是，Del-Rips 在持久图中没有像 Alpha 那样捕捉到第二个 H_1 类，而它捕捉到的唯一 H_1 类，其持续时间也不及 Alpha 中对应的 H_1 类。

补充说明一下，读者可能会想知道，持久图（PD）中是否应该捕捉到第二个环。尽管存在某个尺度值，此时球体会以一种揭示数据中两个环的方式重叠，但在具体应用中，捕捉持久性较弱的那个环是否有价值尚不明确。使用 Alpha 肯定能很好地捕捉到另一个环，但代价是计算效率较低，因为它需要适当地平衡不同维度单形上的权重。第 4 节将部分探讨，我们能否牺牲 Alpha 对数据底层拓扑结构的保真度，以换取计算速度的提升。

3.2 实现与运行时间分析

这里展示了使用德劳内-Rips 过滤在不同数据集上计算持久图的性能实证结果。计算持久同调群时，我们将系数域固定为 $\mathbb{Z}_2$ 。我们用 Python [Misb], 实现了算法 1，该算法会在不同尺度下构造德劳内-Rips 过滤。随后，将此过滤传递给持久同调算法工具箱（PHAT）[Bau+17] ，以构造边界矩阵、约简边界矩阵并提取持久对。实验在一台配备 Intel i7-10875H 处理器（运行频率为 2.3 赫兹）、64 GB 内存并运行 Ubuntu 20.04.5 LTS 的计算机上进行。运行时间以将数据集输入相应算法到生成持久图之间的耗时来衡量。作为比较，我们选用了 Python 模块 Ripser [TSB18] 和 Python 软件包 Cechmate[TS]中的 Alpha 实现。图 2 展示了从带噪声的 2-球面中采样的点数变化时，运行时间如何变化。图中每个数据点表示 10 次试验的中位数。每个数据点上的箱线图显示了运行时间的最大值、最小值和四分位距。同样，图 3 展示了球面维度变化时的运行时间。

算法 1：计算 Delaunay-Rips 过滤的算法

```
输入： P = (p1, . . . , pn)， dim = 要计算的最大同调维度
输出：Delaunay-Rips 过滤
三角剖分 ← Delaunay(P)
过滤 ← [([i],0)] 对 1 ≤ i ≤ n /* 将 0-单形添加到过滤中 */
对每个单纯形 ∈ 三角剖分执行
    对 d ← 1 到 dim + 1 执行
        faces ← 当前 simplex的所有 d-单纯形子集
        对于每个面 ∈ 面 执行
            如果面 ∉ 过滤且 d == 1 则
                value ← 距离(face[0], face[1]) /* 计算欧几里得距离 */
                过滤.append((face, value)) /* 添加 1-单形 */
            end
            如果面 ∉ 过滤且 dim > 1 ，则
                找到具有最大 value 的 subface 中的 face
                filtration.append((subface, value)) /* 添加高阶单形 */
            end
        end
    end
end
```

请注意，Rips 计算在更高维度下会变慢，因为我们要求 Ripser 计算单形，直至数据集的环境维数，与 Alpha 和德劳内-Rips 过滤的计算方式相同。例如，对于位于 3-球面上的数据点（在 $\mathbb{R}^4$ 中），Rips、Alpha 和德劳内-Rips 会计算 H_0、H_1、H_2 和 H_3 类。特别是对于 Rips，维度的增加会使单形的数量呈指数级增长，因此

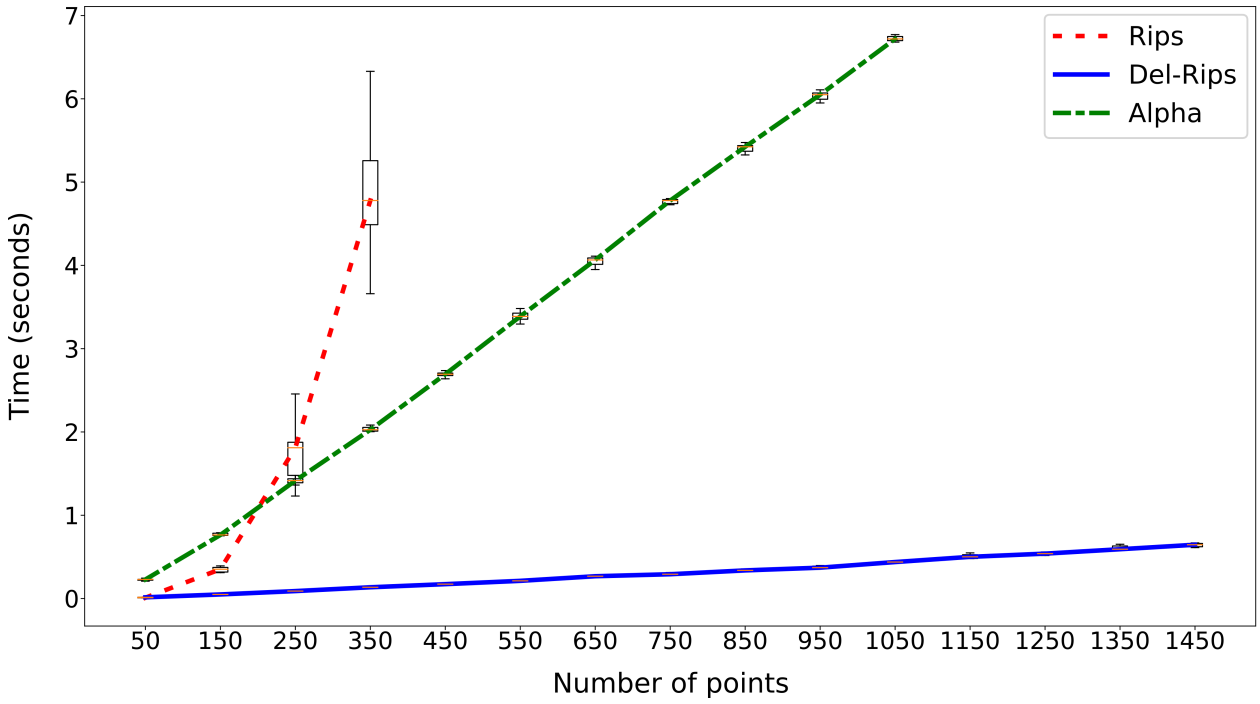


图 2：随着点数增加，Rips、Alpha 和 Delaunay-Rips 的运行时间比较。数据集取自半径为 1、含有 0.1 噪声的 2-球面表面。允许运行的最长时间为 7 秒。

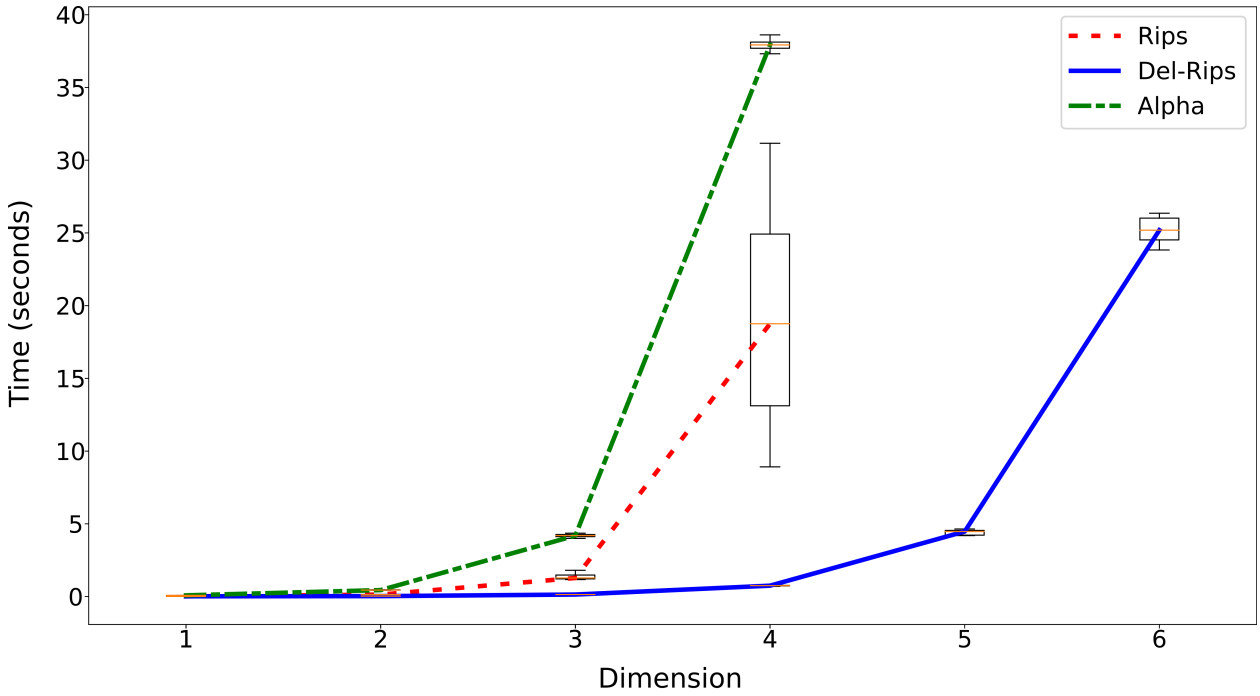


图 3：随着 d-球面维度增加，Rips、Alpha 和 Delaunay-Rips 的运行时间比较。（注意，x 轴上的 1、2、3 分别对应 1-球面、2-球面和 3-球面。）数据集为从半径为 1、含有 0.1 噪声的 d-球面表面采集的 100 个点。允许运行的最长时间为 45 秒。

增大边界矩阵的规模，从而导致计算速度变慢。

请注意，尽管 Alpha 计算和德劳内-Rips 都依赖于计算 Delaunay三角剖分，但与德劳内-Rips 相比，Alpha 计算的速度会显著变慢。这是因为计算多个沃罗诺伊单元相交时对应的尺度，以便分配单形权重，会带来额外的计算开销。由于德劳内-Rips 避免了这项计算（在 Cechmate 中 Alpha 的 Python 实现里，这项计算开销很大），因此在实际应用中能显著加速。需要注意的是，Alpha 的其他实现（例如 GUDHI [Rou23]）可能会比德劳内-Rips 的 Python 实现更快，这取决于实现细节（例如所选的编程语言）。尽管 Alpha 和 Del-Rips 构建的过滤复形相同，但我们预计，得益于这些方法优化版本中更快的单纯形尺度计算，德劳内-Rips 在实际应用中会有一定优势。

3.3 Delaunay-Rips复形的稳定性性质

为了理解对底层点云数据的扰动对所得德劳内-Rips 持久图的影响，我们采用 [Wel97] 对点云的 ε-扰动的定义。若 $P = \{p_1, \ldots, p_n\} \subset \mathbb{R}^D$ 表示一个点云，我们用 $P' = \{p'_1, \ldots, p'_n\}$ 表示 P 的一次扰动：可以想象在每个点周围放置半径为 $\varepsilon < \left(\min_{p_i,\ p_j \in P}\{d(p_i, p_j)\}\right)/2$ 的球，并从每个球中选取一个扰动后的点 $p'_i \in P'$。

定义 3.2. 如果对于每个 $p \in P$，恰好存在一个 $p' \in P'$，使得 $d(p, p') < \varepsilon$，并且反过来，p 是 P 中与 p' 的距离在 ε 以内的唯一点，则称 $P' = \{p'_1, \ldots, p'_n\}$ 是 $P = \{p_1, \ldots, p_n\}$ 的一个 ε-扰动。我们将每一对 p、p' 称为一个扰动对。

换言之，存在一个双射 $\rho\colon\ P \to P'$，它将每个点 $p_i \in P$ 与一个点 $p'_i \in P'$ 关联起来，使得 $d(p_i, \rho(p_i)) < \varepsilon$。

引理 3.3. 已知 P' 是某个 $P \subset \mathbb{R}^D$ 的 ε-扰动，则存在一个扰动对 $p_i \in P$ 和 $p'_i \in P'$，使得

$$d_H(P, P') = d(p_i, p'_i).$$

证明。首先，在所有扰动对中找出彼此距离最远的扰动对 $x \in P$ 和 $x' \in P'$，并令

$$\delta := d(x, x') = \max_{p \in P, p' \in P'} \left(d(p, p')\right). \tag{1}$$

通过这样的选择，我们确保对于任意 $p' \in P'$，必有 $p' \in P_\delta$，因为 $d(p, p') \le d(x, x') = \delta$。类似地，对于任意 $p \in P$，$d(p, p') \le d(x, x') = \delta$，因此 $P \subseteq P'_\delta$。所以，根据定义 2.13，$d_H(P, P') \le \delta$。现在，假设 $\lambda < \delta$。对于任意 $p' \in P'$

$$d(x, p') \ge \min_{p' \in P'} \left(d(x, p')\right) = d(x, x') = \delta > \lambda.$$

因此，$x \notin P'_\lambda$，因为它不在 P' 中任何点周围的 λ-球内，所以对于每个 $\lambda < \delta$，$d_H(P, P') > \lambda$。因此，$d_H(P, P') = \delta = d(x, x')$。也就是说，$P$ 与 P' 之间的豪斯多夫距离，恰好等于在 ε-扰动中，P 中某个点发生扰动的最大距离。 □

为了利用定理 2.12，我们将探寻能够确保点受到扰动时底层 Delaunay三角剖分不发生变化的条件。若有限的 $P \subset \mathbb{R}^D$ 与 ε-扰动 P' 关于扰动配对具有相同的 Delaunay三角剖分，则称二者满足此性质，当且仅当对于所有 $0 \le k \le D$，$\sigma = [p_{i_0} \ldots p_{i_k}] \in$Del$(P)$ 等价于 $\sigma' = [p'_{i_0} \ldots p'_{i_k}] \in$Del$(P')$；其中，$p'_{i_l}$ 是 P' 中与 p_{i_l} 唯一对应的点，且 $d(p_{i_l}, p'_{i_l}) < \varepsilon$。作为抽象单纯复形，Del$(P)$ 和 Del(P') 无法区分，因为扰动对的对应关系在单形之间建立了一一对应。因此，我们不会区分这些复形中相对应的单形。

定理 3.4. 设 $P \subset \mathbb{R}^D$ 为一个点云，P' 是 P 的 ε-扰动，并且相对于扰动配对具有相同的Delaunay三角剖分；称其为 K。设 f_P、$f_{P'}\colon\ K \to \mathbb{R}$ 为单调函数，它们分别通过将Delaunay-Rips尺度赋予 K 中的单纯形（在 P 和 P'中观察）来定义。对于每个维度 p，持久图之间的瓶颈距离不超过点云 P 与 P'之间豪斯多夫距离的两倍：

$$W_\infty\left(Dgm_p(f_P), Dgm_p(f_{P'})\right) \le 2d_H(P, P').$$

证明. 我们将利用定理 2.12的结果。由于 P 是有限点云，因此存在一个单纯形$\sigma \in K$ ，使得

$$||f_P - f_{P'}||_\infty = |f_P(\sigma) - f_{P'}(\sigma)|.$$

由于单纯形的 Delaunay-Rips 尺度由该单纯形中距离最远的两个点决定，因此存在 $p, q \in P$和 $r', s' \in P'$，它们都是 σ中的 0-单纯形，使得 $f_P(\sigma) = d(p, q)$，且 $f_{P'}(\sigma) = d(r', s')$。 p, q和 r', s'可能构成两个扰动对（例如， $p' = r'$和 $q' = s'$），但并非总是如此。例如，决定 Del(P)中 σ尺度的点 p 和 q 经过扰动后，可能移动到 p'和 q'（它们必然都属于 Del(P')中的 σ），且彼此比决定 Del(P')中 σ 尺度的点 r' 和 s'更近。我们分别考虑这两种情况。

情形 1： $p' = r'$ 且 $q' = s'$。不失一般性，假设

$$d(p, q) > d(p', q').$$

Then

$$\begin{aligned}
|f_P(\sigma) - f_{P'}(\sigma)| &= |d(p, q) - d(r', s')| \\
&= |d(p, q) - d(p', q')| \\
&= d(p, q) - d(p', q') \\
&\leq d(p, p') + d(p', q') + d(q', q) - d(q', p') \\
&= d(p, p') + d(q', q) \\
&\leq d(x, x') + d(x, x') \\
&= 2d_H(P, P'),
\end{aligned}$$

其中， $x \in P$ 和 $x' \in P'$ 是引理 3.3 中 P 和 P' 内彼此距离最远的配对点。情形 2： $p' \neq r'$ 或 $q' \neq s'$。如前所述，

$$d(r', s') \geq d(p', q') \tag{2}$$

$$d(p, q) \geq d(r, s), \tag{3}$$

根据假设， σ 中的哪一对顶点决定了 Del(P) 和 Del(P') 中的尺度。现在，有两种子情形需要考虑。

情形 2a： $d(p, q) \geq d(r', s')$。利用公式 2，可得

$$|f_P(\sigma) - f_{P'}(\sigma)| = |d(p, q) - d(r', s')| = d(p, q) - d(r', s') \leq d(p, q) - d(p', q')$$

其余部分可由与情形 1 相同的论证得出。

情形 2b： $d(p, q) < d(r', s')$。利用公式 3，可得

$$|f_P(\sigma) - f_{P'}(\sigma)| = |d(p, q) - d(r', s')| = d(r', s') - d(p, q) \leq d(r', s') - d(r, s)$$

其余部分在适当重新标记后，可按情形 1 的方法推得。

最后，应用定理 2.12，我们得出，与 P 和 P' 分别关联的持久图之间的瓶颈距离不超过这两个点云之间豪斯多夫距离的两倍：

$$W_\infty \left(\mathrm{Dgm}_p(f_P), \mathrm{Dgm}_p(f_{P'})\right) \leq ||f_P - f_{P'}||_\infty \leq 2d_H(P, P').$$

□

我们看到，当点云上的底层 Delaunay三角剖分保持不变时（相对于扰动配对而言），使用 DR算法为单形指定尺度可以保证相应持久图的稳定性。尽管已知对于处于一般位置的点云 P，存在一个足够小的 $\varepsilon > 0$，使得每个 ε-扰动 P' 都会具有相同的 Delaunay三角剖分[BDG13], ，但 ε 的大小可能非常小，这使人们对上述稳定性结果在实际应用中的实用性产生疑问，因为实际应用中存在测量不确定性。下一节将探讨当扰动数据点导致底层 Delaunay三角剖分发生变化时可能出现的情况。

3.4 持久图不稳定性

DR构造以稳定性为代价获得了计算效率。我们通过一个简单而清晰的例子，说明在底层数据发生扰动时，数据到图的变换可能出现不连续性。在图 4 中，从左向右看，我们设想将最右侧的点（红色）向右移动，朝向另外三个点的唯一内切圆。当最右侧的点位于圆内时，存在一个具有非零持久性的 H_1 类。当这 4 个点位于同一个圆上时，该类会立即消失。非正式地说，这意味着在前两个阶段，DR 将这 4 个点视为形成了一个环；而当这些点变为共圆时，DR 会立即失去对该环的识别。在图中，我们标出了这些点的Delaunay三角剖分，以展示发生边翻转时最右侧（红色）点的位置（即四个点位于同一个圆上时）。我们看到的是，在视觉结构上非常相似的点云，却产生了截然不同的持久图。将最右侧的点向圆内进行任意小的扰动，会得到与向圆外进行任意小的扰动截然不同的持久图。下面我们将正式证明与这一特定点配置相关的持久图的不稳定性。

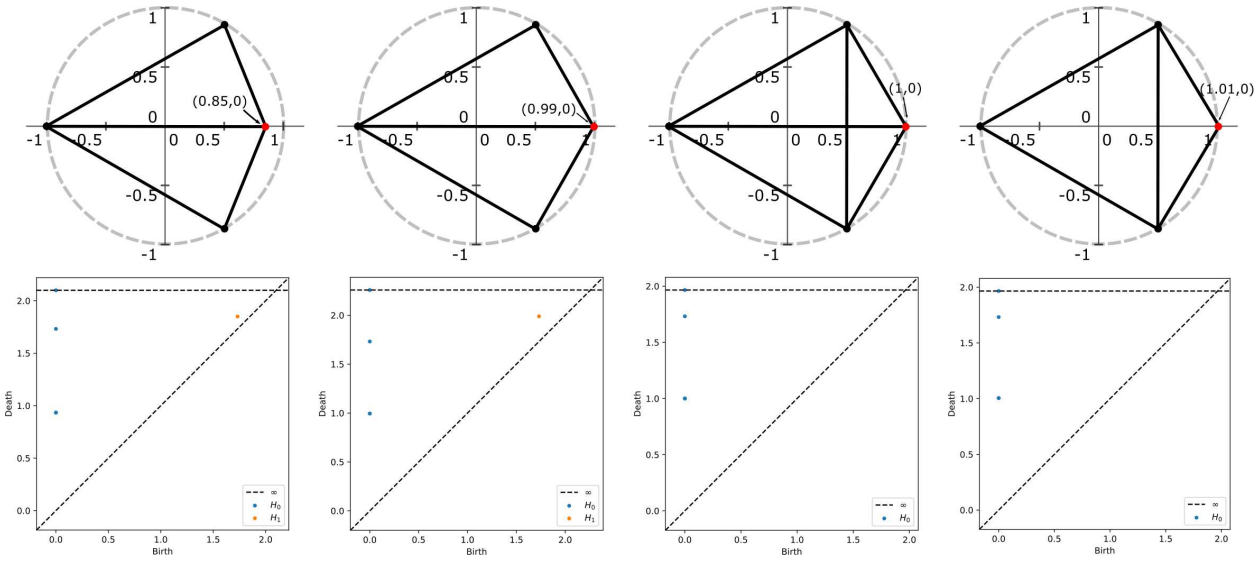


图 4：4 点示例的持久图，其中最右侧的点沿 x 轴水平向右移动。请注意，在前两个阶段（从左数），图中存在一个具有非零持久性的 H_1 类。 H_1 类在最后两个阶段消失。

引理 3.5。设 $P' = \{(-1, 0), (1, \frac{\sqrt{3}}{2}), (1, -\frac{\sqrt{3}}{2}), (1-x, 0)\}$，其中 $0 < x < \delta < 2-\sqrt{3}$。使用 Delaunay-Rips 复形在此点云上构造过滤后，只有一个具有非零持久性的 H_1 同调类。

证明。我们有 $P' = \{(-1, 0), (1, \frac{\sqrt{3}}{2}), (1, -\frac{\sqrt{3}}{2}), (1-x, 0)\}$，其中 $0 < x < \delta < 2-\sqrt{3}$。我们的过滤有 4 个关键尺度值，$t = 0 < \sqrt{1-x+x^2} < \sqrt{3} < 2-\delta$ 如图 5 所示。

我们构造一个边界矩阵 B，其元素取自域 $\mathbb{Z}_2$，并使用 [EH10]，第VII章中的标准约化算法将其约化为 $\overline{B}$。该算法计算单纯形配对：它们分别使 H_0 和 H_1 持久图中的持久同调类产生并消亡。B 的列顺序由 DR 过滤所给出的单纯形顺序决定，同时确保每个单纯形都排在其各个面之后。

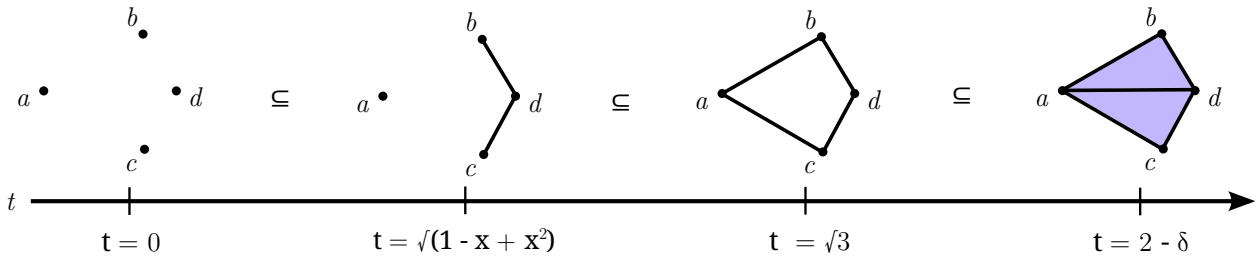


图5：包含 4 个关键尺度值的过滤，其中 $0 < x < \delta < 2 - \sqrt{3}$.

$$B = \begin{array}{c} \\ a \\ b \\ c \\ d \\ bd \\ cd \\ ab \\ ac \\ ad \\ abd \\ acd \end{array} \begin{array}{c} \begin{array}{cccccccccc} a & b & c & d & bd & cd & ab & ac & ad & abd & acd \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

$$\overline{B} = \begin{array}{c} \\ a \\ b \\ c \\ d \\ bd \\ cd \\ ab \\ ac \\ ad \\ abd \\ acd \end{array} \begin{array}{c} \begin{array}{cccccccccc} a & b & c & d & bd & cd & ab & ac & ad & abd & acd \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}.$$

通过计算每个单纯形的尺度，可以得到 H_0 类的持久对及其持久图坐标（出生/死亡对）：

$$(a, N/A) : (0, \infty)$$
$$(b, ab) : (0, \sqrt{3})$$
$$(c, cd) : (0, \sqrt{1 - x + x^2})$$
$$(d, bd) : (0, \sqrt{1 - x + x^2}).$$

同样， H_1 持久对为

$$(ad, abd) : (2 - x, 2 - x)$$

and

$$(ac, acd) : (\sqrt{3}, 2 - x).$$

唯一具有非零持久性的 H_1 类是（$\sqrt{3}, 2 - x$）。 □

定理 3.6。设 $(\mathcal{P}, d_H)$ 为配备豪斯多夫度量的点云空间，$(\mathcal{D}, W_\infty)$ 为配备瓶颈度量的持久图空间。令

$$Pers_1 : \mathcal{P} \to \mathcal{D}$$

其中，$\text{Pers}_1(P)$ 是使用 Delaunay-Rips复形构造的点云 P 的 H_1 类持久图。此映射不连续。

证明。令 $P \in \mathcal{P}$ 为 $P = \{(-1, 0), (\frac{1}{2}, \frac{\sqrt{3}}{2}), (\frac{1}{2}, -\frac{\sqrt{3}}{2}), (1, 0)\}$.。注意，所有点都位于单位圆上，因此德劳内 1-骨架中任意两点之间都有一条边（见图 4 的第三帧）。对于这一点配置，连接点 $(\frac{1}{2}, \frac{\sqrt{3}}{2})$ 和 $(\frac{1}{2}, -\frac{\sqrt{3}}{2})$ 的竖直边与由全部四个点构成的圈在完全相同的尺度值出现。因此，在 DR 过滤中，边界由四条外侧边构成的 H_1 类在诞生的同时便消亡。

固定 $\varepsilon = 0.1$。现在我们证明，对于任意 $\delta > 0$，存在 $P' \in \mathcal{P}$，使得 $d_H(P, P') < \delta$，但 $W_\infty(\text{Pers}_1(P), \text{Pers}_1(P')) \geq \varepsilon$.取 $P' = \{(-1, 0), (\frac{1}{2}, \frac{\sqrt{3}}{2}), (\frac{1}{2}, -\frac{\sqrt{3}}{2}), (1 - x, 0)\}$，其中 $0 < x < \min\{\delta, \frac{2-\sqrt{3}}{2}\}$。这是对 P 的一次小扰动，具体做法是将点 (1,0) 推入单位圆内部，从而使这些点处于一般位置（示例见图 4 的第二帧）。直接计算可得，P 与 P' 之间的豪斯多夫距离 d_H 为

$$d_H(P, P') = x < \delta.$$

回忆一下，$\text{Pers}_1(P)$ 没有持久性非零的 H_1 类。因此，为了计算 $W_\infty(\text{Pers}_1(P), \text{Pers}_1(P'))$，我们必须将 $\text{Pers}_1(P')$ 的 H_1 类与对角线匹配。根据引理 3.5 计算可得，$\text{Pers}_1(P')$ 的 H_1 类的出生时间为 $\sqrt{3}$，死亡时间为 $2 - x$。使用最大范数，我们得到

$$W_\infty(\text{Pers}_1(P), \text{Pers}_1(P')) = (2 - x) - (\sqrt{3}) \geq 2 - \frac{2 - \sqrt{3}}{2} - \sqrt{3} \geq 0.1 = \varepsilon.$$

因此，Pers_1 在 P 处不连续。 □

顺便指出，点云空间中任何被豪斯多夫距离从上界定的度量都会产生这种不连续性（例如格罗莫夫-豪斯多夫距离）。这让我们得以了解持久图的德劳内-Rips构造何时可能出现不稳定性——也就是说，点不处于一般位置时。

到目前为止，我们已从数学上证明：当点的位置受到扰动时，若数据集的 Delaunay 三角剖分保持不变，则其持久图具有理论稳定性。然而，我们不应期待现实应用中总能满足保证稳定性的条件；本节还表明，当底层的 Delaunay 三角剖分确实发生变化时，图之间的变化程度可能无法由底层数据的变化程度所控制。因此，我们不禁要问，这种不稳定性在实践中究竟有多大影响?

4 使用 Rips、Alpha 和 Delaunay-Rips 过滤的机器学习模型性能

尽管存在一些特殊情况，可能会出现持久图不稳定的现象（如第 3.4 节所示），但我们在此说明，这种不稳定性在应用中可能影响不大。我们展示了 DR 在合成数据和真实数据场景中的机器学习鲁棒性。对于合成数据，我们在多类分类任务中使用随机森林分类器。对于真实数据，我们使用线性核支持向量机训练二元分类模型。

4.1 SyntheticShape 数据的分类

为了测试 DR 对点云数据中点的位置扰动程度变化的鲁棒性，我们使用 Rips、Alpha 和 DR 过滤，在通过对不同流形进行随机采样并添加随机噪声以扰动点的位置而生成的点云上，构建机器学习分类模型。虽然在底层 Delaunay三角剖分保持不变时，使用 DR 过滤生成的持久图具有稳定性（见第 3.3 节），但实际上，即便噪声程度较低，点云的底层 Delaunay三角剖分也可能发生变化。

图 6 展示了我们实验的一般流程。我们针对 6 种形状类别（圆、球面、环面、随机、三个簇，以及三个簇中各自包含三个簇）分别生成了 100 个点云，每个点云包含 500 个数据点，且均为 $\mathbb{R}^3$ 的子集。对于固定的噪声水平 ν，我们使用从环境空间中随机选取、最大模长为 ν 的向量扰动每个点云中的点。对于每个点云，我们使用 Rips、Alpha 和德劳内-Rips 过滤计算 0、1 和 2-持久性图，然后利用 scikit-tda 库中 Python 包“persim”提供的持久图像（PIs）将所得的图向量化 [Sau+]。随后，我们使用 scikit-learn 中的实现训练随机森林分类器，将所得特征向量作为输入 [Ped+11]。为了评估和比较不同的过滤方法，我们使用 10 折交叉验证，计算训练好的模型在留出数据上的分类准确率中位数。

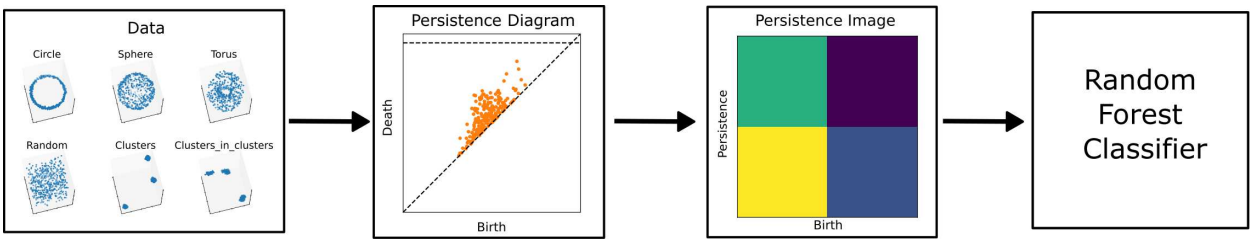


图 6：第一步是根据 6 个形状类别生成数据点。下一步是为每个数据集计算每个维度（最高至 3 维）的持久图（上图是 3 维中 500 个随机点的 H_1 持久图）。然后，将这些图转换为具有 2×2分辨率网格的持久图像。最后，我们将图像展平为向量，并训练随机森林分类器，以区分这 6 个形状类别。

固定一个同调类以及点云扰动的噪声水平 ν 后，我们确定所有过滤方法和所有样本生成的持久对的出生范围和持久性范围，并将该区域分割为具有 2×2 分辨率的持久图像。例如，我们遍历使用 $\nu = 0.20$生成的三种过滤方法的所有 H_1持久图。然后，我们找出最大的出生范围和持久性范围，并使用这些值设置 2×2分辨率网格，以生成持久图像。这样做的目的是确保在比较对应于不同过滤方法的持久图像像素时能够进行公平比较，因为对应的持久对会落入对应的像素中（在比较特征重要性时，我们会在图 8 中利用这一设计）。

图 7 展示了我们的模型在噪声水平不断增加时的中位数准确率（我们绘制了箱线图，以展示 10 折交叉验证所得准确率的分布）。作为基线比较，请注意，如果我们的机器学习模型对测试数据进行随机分类，预计准确率为 $1/6 \approx 17\%$。由于在每个噪声水平下，我们观察到的中位数准确率都超过 70%，因此我们的模型确实找到了 6 个形状类别之间具有区分性的特征。

值得注意的是，如图 7 所示，随着噪声水平增加，机器学习模型性能的下降在德劳内-Rips、Alpha 和 Rips 过滤之间非常相似。我们确实注意到，在所有噪声水平下，使用 Alpha 过滤生成持久图时的中位数分类准确率都略高于其他方法。然而，除单个噪声水平（.15）外，所有中位数准确率都处于另外两种方法的范围之内。

除了模型准确率之外，一个重要的考量因素是模型使用了哪些特征来实现这一准确率。识别显著特征有助于提升模型可解释性 [RSG16], ，也可能引导建模者采用降低输入维度的方法，从而进一步提升模型性能。为了进一步比较三种过滤方法，我们比较了训练过程中学得的最重要拓扑特征

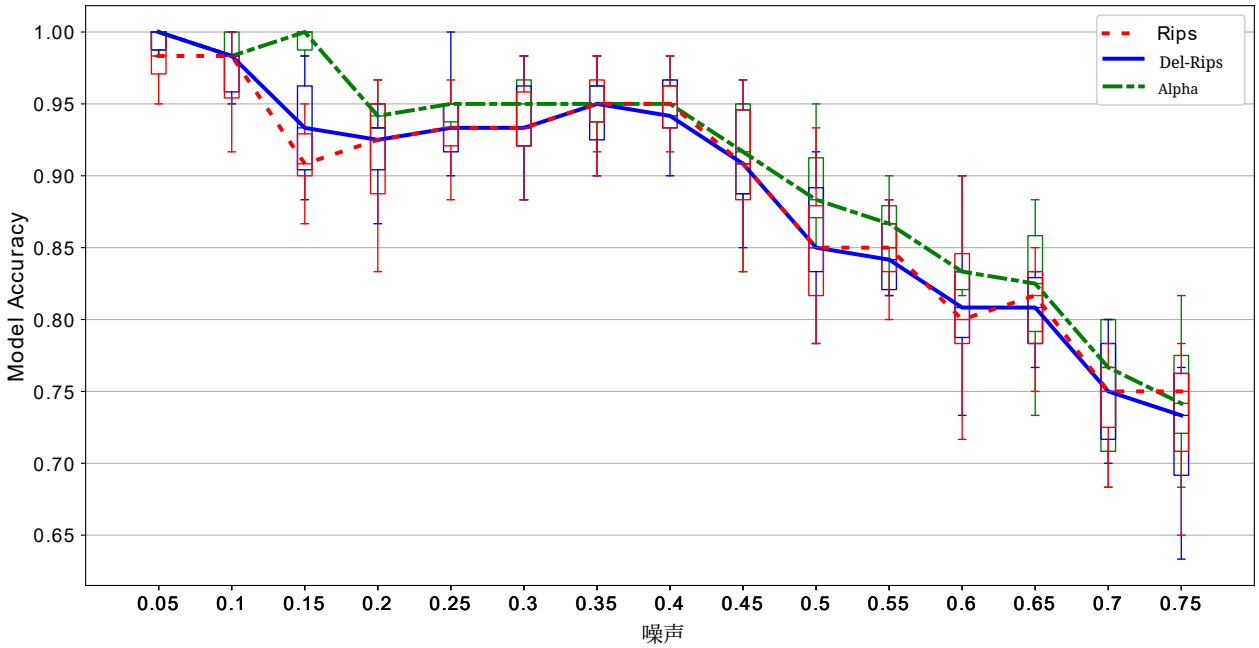


图 7：针对原始点云，在不同噪声水平下，基于 Rips、德劳内-Rips 和 Alpha 持久图计算得到的随机森林分类准确率中位数图。每个噪声水平对应的箱线图展示了 10 折交叉验证中的准确率范围。

，这些特征由随机森林分类器确定。我们从$\nu = 0.20$ 噪声数据生成的持久图开始。我们为每个持久图生成了以下分辨率的持久图像：

- H_0 图分辨率： 5×1
- H_1 图分辨率： 5×5
- H_2 图分辨率： 5×5

因此，我们为 600 个样本中的每个样本（6 个类别各 100 个样本）生成了 55 维特征向量。我们使用与之前相同的参数训练了随机森林分类器，并采用 70-30 的训练集-测试集划分；随后使用内置的评估功能 [Ped+11]，通过基尼重要性量化特征重要性。特征的基尼重要性是根据该特征带来的基尼指数[KA17] 下降幅度计算的。因此，基尼重要性越高，该特征在分类决策中的作用就越大。图 8 展示了表示特征重要性的热力图。请注意，对于我们用于生成持久图的不同过滤，生成的持久图像中都有相似的像素被机器学习模型判定为重要特征。例如，在基于德劳内-Rips 的随机森林分类器所用的 $5+25+25$像素特征空间中，最重要的特征是左下方（深蓝色）的 H_2 像素（基尼重要性为 0.19）和底部的 H_0 像素（基尼重要性为 0.17）。这些区域与基于 Alpha 和 Rips 的分类器所判定的最重要特征所在区域相似。因此，我们有理由认为，无论采用何种过滤方法，机器学习模型都学会了识别出能够区分这些形状的相似特征。

我们的数据分析代码已在在线代码库 [Misa]中附有完善文档，以确保结果可复现。

4.2 SleepState分类

接下来，我们研究 DR 在涉及生物物理数据的分类问题中的适用性。我们的目标是展示，在计算效率可能构成相关限制的问题中，DR 与 Alpha和Rips滤子在模型性能指标方面具有相当的有效性。我们的应用是重新分析 [Chu+21] 所采用的数据和方法；该研究的作者开发了一种机器学习模型



图 8： H_0、 H_1 和 H_2 持久图 PI向量化的热力图。我们基于每种过滤函数选取 55 个特征来训练随机森林分类器，并在相应的分类器模型中评估这些特征的重要性。每个像素中的数值表示该像素在随机森林分类决策中的基尼重要性。数值越高，该像素越重要。

，利用通过心电图（EKG）采集的瞬时心率（IHR）时间序列，对参与者的睡眠阶段进行分类。图 9 展示了我们所采用的数据到模型开发流程的高层次可视化。需要说明的是，本文（研究 DR 过滤的性质和用途）的作者并未直接处理这些数据；相反，经过处理的数据是在获得 [Chu+21] 的作者许可后取得的。

[Chu+21] 的作者训练了一个机器学习模型来预测睡眠状态，该模型基于高维点云拓扑特征的统计量进行训练，而这些点云由 IHR 时间序列的延迟嵌入构建。这种方法基于这样一种理念：正如 Takens 的著名定理所示，观测到的时间序列的延迟嵌入可以恢复未知状态空间中底层吸引子的动力学特征 [Tak81]。这些吸引子的拓扑/几何特征（以持久图表示）可能有助于区分不同的动态状态。基于类似原理的方法也已应用于三维动作捕捉数据等领域 [VRT16]。

用于训练和测试模型的数据包括来自长庚纪念医院（CGMH）的 90 条心电图记录，构成 CGMH 训练数据库。最初使用这些数据的论文是[MLW18] ，其中对数据和数据采集作了更详细的介绍。

每条记录的采样频率为 200 赫兹，每个 30 秒时段均由人工标注为“清醒”或“睡眠”状态。睡眠状态进一步分为 5 种睡眠阶段之一：第 1 至第 4 阶段或快速眼动睡眠，因此可形成一个 6 类分类问题。对每条记录应用标准的 R 峰检测算法，并使用 5 次心搏中值滤波器去除伪迹（做法与 [XS01]类似）。基于这些预处理数据，以 4 赫兹的频率计算 IHR 时间序列。为提高效率，我们将 IHR 时间序列延迟嵌入至 $\mathbb{R}^3$，并将连续坐标之间的延迟参数设为 5（即，延迟向量由点（x_i、 x_{i+5}、 x_{i+10}） $\in \mathbb{R}^3$构成，取时间序列中每隔 5 个点作为嵌入点的坐标）。将连续延迟向量起始坐标之间的索引数（即步长）设为 1，以确保所有时间点至少包含在一个延迟向量中。

将每个时段嵌入到 $\mathbb{R}^3$ 后，我们使用嵌入点云的德劳内-Rips、Rips 和 Alpha 过滤，计算了 H_0、 H_1 和 H_2 持久图。按照 [Chu+21]的方法，通过计算图中持久对的样本统计量，将所得持久图转换为固定长度的特征向量，具体如下：

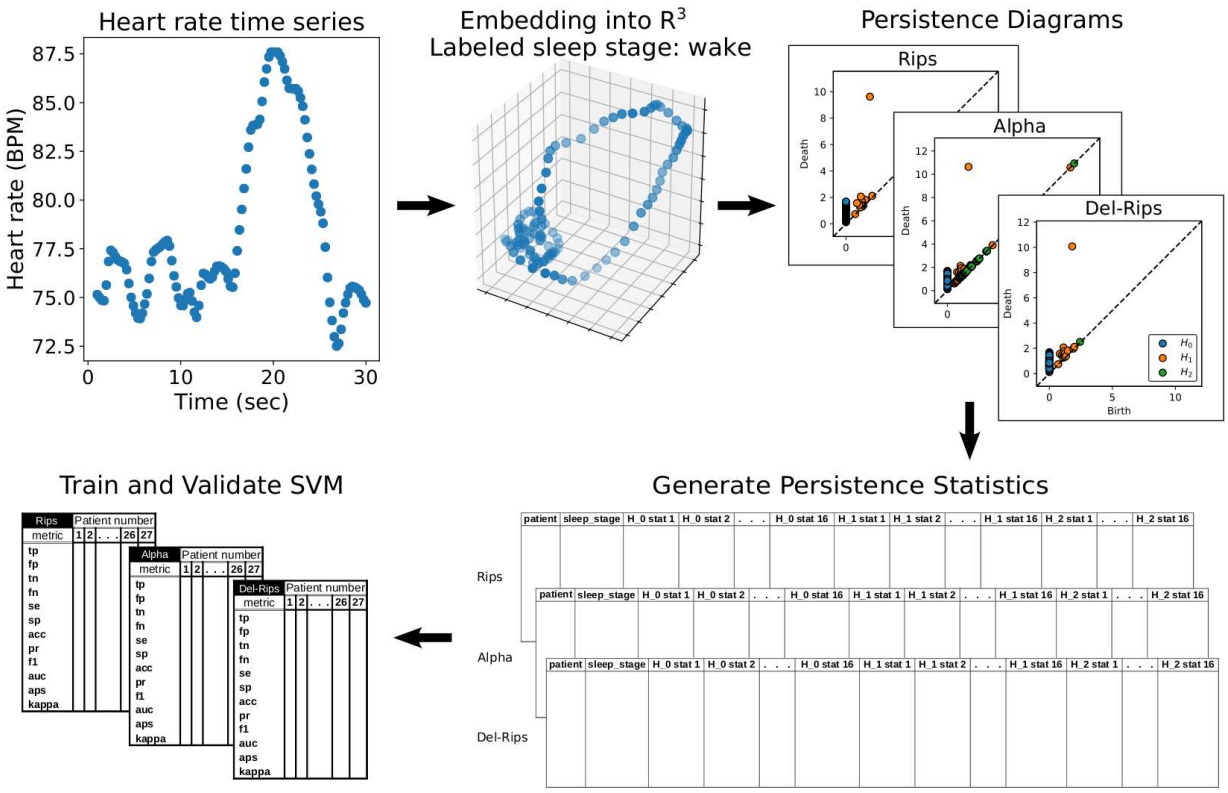


图 9：所用数据到模型流程图。从左上方开始，展示一个 30 秒清醒时段及其延迟嵌入到 $\mathbb{R}^3$ 的示例、根据点云上的德劳内-Rips、Alpha 和 Rips 过滤计算出的相应持久图，最后是使用持久性统计量进行后续向量化，并将其用于训练和验证支持向量机二元分类模型。

对于每个同调维数 $p = 0$、1、2，我们构建两组数据： M_p，即所有持久对的均值集合；以及 L_p，即每对的持久性（死亡时间减去出生时间）。对于每组数据，我们计算均值、标准差、偏度、峰度、第 25 百分位数、第 50 百分位数、第 75 百分位数，以及持久熵 [Chi+15]。每个维度的 p持久图对应 16 个持久性统计量。对于空的 H_p 图，赋值为 NaN；随后从训练集中剔除这些样本。总计为每个样本计算 48 个持久性统计量（维度 $p = 0$、1、2 各 16 个），由此得到一个 48 维特征空间，并在其上训练支持向量机模型。

总共使用了 67188 个 48 维拓扑特征向量（每个 30 秒时段对应一个），来训练一个采用线性核并设置平衡类别权重的支持向量机，以应对睡眠样本和清醒样本数量不平衡的问题。总共训练了 3 个模型，每种过滤方法对应一个。随后，使用 CGMH验证集中的 27 名留出参与者对每个训练好的模型进行验证：针对每位验证参与者的时段，计算灵敏度 (se)、特异度 (sp)、准确率 (acc)、精确率 (pr)、F1分数 (f1)、AUC分数 (auc)、平均精确率 (aps) 和 Kappa系数 (kappa)。为评估德劳内-Rips 过滤与 Rips 或 Alpha 过滤在性能指标方面的差异是否具有统计显著性，我们使用 SciPy [Vir+20] 中的 Mood中位数检验，检验 27 名验证参与者各项性能指标的分布，并分别比较 Rips 与德劳内-Rips 过滤、Alpha 与德劳内-Rips 过滤。在此背景下，Mood中位数检验用于检验各项性能指标的中位数相同这一零假设。

我们的模型构建与原论文中的模型有几个关键区别。首先， [Chu+21] 的作者从 IHR 时间序列的高维（120 维）延迟嵌入中提取了拓扑特征。这种高维嵌入给 Alpha 和德劳内-Rips 过滤带来了挑战，因为二者都需要计算 Delaunay三角剖分，而这类剖分难以高效扩展到高维空间。

(a)	Rips		Alpha		Del-Rips	
	中位数	iqr	中位数	iqr	中位数	iqr
se	0.565217	0.428794	0.569892	0.380411	0.555556	0.433198
sp	0.925065	0.271714	0.917511	0.270439	0.917271	0.272863
acc	0.852941	0.161052	0.848276	0.154095	0.852573	0.161564
pr	0.439636	0.316970	0.448718	0.344648	0.424307	0.301726
f1	0.517986	0.140499	0.489655	0.136024	0.472727	0.170183
auc	0.866199	0.083241	0.864330	0.079358	0.866646	0.096021
aps	0.594109	0.180105	0.555916	0.160430	0.552859	0.159507
卡帕系数	0.393895	0.178491	0.368380	0.197346	0.356587	0.192436

(b)	Rips与Del-Rips的p值	Alpha与Del-Rips的p值
se	1.000000	1.0
sp	1.000000	1.0
acc	1.000000	1.0
pr	1.000000	1.0
f1	0.276303	1.0
auc	1.000000	1.0
aps	0.102470	1.0
卡帕系数	1.000000	1.0

表 1：(a) 由 3 种分类器分别基于 DR、Rips 或 Alpha 过滤所确定的拓扑特征进行训练后，在 27 名验证参与者中，各分类模型性能指标的中位数和四分位距。(b) 将 DR 模型性能指标与 Rips 和 Alpha 模型性能指标进行比较后，Mood中位数检验得到的 p 值。

此外，原模型仅使用从 H_0 和 H_1 图中得出的统计量进行训练，而我们还纳入了 H_2 图。我们指出，与 [Chu+21]报告的结果相比，这些改变使模型性能略有下降；不过，我们的目标是评估不同过滤之间的差异，而非改进先前发表的分类模型。

表 1 展示了各模型性能指标在验证参与者中的中位数和四分位距，以及 Mood中位数检验的 p 值。我们观察到，三种过滤方法的性能指标非常相似；根据不同指标，DR、Alpha 或 Rips 的中位数表现各有最佳。对于 Mood中位数检验，较小的 p 值表明两个分类模型在相应性能指标的中位数方面存在显著差异。请注意，表 1 中的 p 值都很大，唯一的例外是比较 Rips 与 DR 时的平均精确率得分（aps）。这表明，考虑到模型在不同验证参与者上的表现差异，我们没有足够证据拒绝“中位数性能指标相同”这一零假设。

我们的数据分析代码在在线代码库 [Misc]中有详尽的文档，便于复现。

5 结论

本文定义并实现了用于点云数据的Delaunay-Rips（德劳内-Rips，DR）过滤，将其在实践中的计算效率与其他基于点云数据构建的标准过滤进行了比较，刻画了其若干稳定性性质，并展示了其在机器学习流水线中应用于各类数据集的情况。我们证明，在点云中各点位置发生足够小扰动的条件下，只要数据集上的底层单纯复形保持不变，持久图便会随数据连续变化。具体而言，我们将持久图之间的瓶颈距离上界限定为原始数据及其扰动对应物之间豪斯多夫距离的两倍。当Delaunay三角剖分因较大扰动而发生变化时，我们通过仔细证明数据度量空间与

图示度量空间之间映射的不连续性，考察了DR过滤的不稳定性。据我们所知，该结果首次使用标准约化算法在符号变量上计算持久图，以服务于形式证明。由于稳定性的理论条件在实际应用中未必总能满足，我们研究了使用DR生成的图中的不稳定性在机器学习流水线中是否会造成显著问题，并发现并非一定如此。

本研究存在若干局限，也有一些值得进一步探究的方向。首先，我们认为，在机器学习建模任务中，用于提取拓扑特征的各过滤方法，其相对表现取决于具体问题。因此，我们无法确定模型表现对过滤方法的不敏感性，以及我们观察到的德劳内-Rips相较于 Rips 和 Alpha 的可比表现，是否能推广到其他情境。

为了保持单形空间固定，我们的稳定性结果是通过要求底层 Delaunay三角剖分保持不变而简化的。尽管第 3.4 节的结果表明，一般而言，底层 Delaunay三角剖分的变化可能导致从数据到图的变换出现不连续性，但目前尚不清楚这些不连续性的程度如何得到更精确的刻画，因此我们只能进行经验评估。

在第 3.2 节中，我们将德劳内-Rips 的 Python 实现与其他过滤方法进行了比较。包括语言选择在内的实现细节，可能会影响德劳内-Rips 相较于其他过滤方法所带来的相对性能提升。若采用 C/C++ 实现德劳内-Rips，基于德劳内-Rips 计算持久图的运行时间，与使用其他复形（Cech、Witness 等）计算持久图的运行时间相比如何？同样，我们的实现（或 C/C++ 实现）在运行时间上，与其他软件包中其他 TDA 方法的实现相比如何（例如 Javaplex [TVA14], Perseus [Nan], Dionysus [Mor], Dipha [Rei], Gudhi [Mar+14]）？最后，在本文撰写期间，软件包 Ripser 发布了一个名为 Ripser++ [SW20] 的更新版本。新方法利用 GPU 加速，对单形处理和持久对查找进行并行化。对于持久图计算，德劳内-Rips 的运行时间与 Ripser++ 相比如何？是否可以利用并行化来提升德劳内-Rips 的计算效率？这些系统性比较不在本研究的范围之内。

完整参考文献

[Tak81] Floris Takens。《检测湍流中的奇异吸引子》。载于：《动力系统与湍流，沃里克 1980》。David Rand 和 Lai-Sang Young 编。柏林、海德堡：Springer Berlin Heidelberg，1981 年，第 366–381 页。 ISBN：978-3-540-38945-3。[Ede93] Herbert Edelsbrunner。《球的并集及其对偶形状》。载于：Discrete & Computational Geometry 13（1993），第 415–440 页。[Wel97] Frank-Uwe Weller。《站点扰动下 Voronoi 邻接关系的稳定性》。载于：CCCG。1997 年。[XS01] Xirong Xu 和 Stephanie Schuckers。《心搏周期数据中伪迹的自动检测》。载于：Journal of electrocardiology 34 Suppl（2001），第 205–10 页。[Hat02] A. Hatcher。《代数拓扑》。《代数拓扑》。剑桥大学出版社，2002 年。 ISBN：9780521795401。 URL：https://books.google.com/books?id=BjKs86kosqgC。[SC04] Vin de Silva 和 Gunnar E. Carlsson。《利用见证复形进行拓扑估计》。载于：基于点的图形学研讨会。2004 年。[CEH05] David Cohen-Steiner、Herbert Edelsbrunner 和 John Harer。《持久图的稳定性》。载于：Discrete & Computational Geometry 37（2005），第 103–120 页。[GO07] Leonidas J. Guibas 和 Steve Oudot。《利用见证复形进行重建》。载于：Discrete & Computational Geometry 40（2007），第 325–356 页。[Lei+08] Gregory Leibon 等。《股票市场网络中的拓扑结构》。载于：美国国家科学院院刊 105.52（2008），第 20589–20594 页。[Mat08] Jiˇrı́ Matouˇsek。《LC 约化产生同构的单纯复形》。载于：Contributions to Discrete Mathematics *[仅电子版]* 3（2008 年 1 月）。

[Sil08] Vin De Silva。《Delaunay三角剖分的一种弱刻画》。载于：Geometriae Dedicata 135.1（2008），第39–64 页。[Cha+09] Fr´ed´eric Chazal 等。《利用持久性为形状构造 Gromov-Hausdorff 稳定签名》。载于：Computer Graphics Forum 28.5（2009），第 1393–1403 页。DOI：10.1111/j.1467-8659.2009.01516.x。电子预印本：https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8659.2009.01516.x。URL：https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8659.2009.01516.x.[CBK09] Moo K Chung、Peter Bubenik 和 Peter T Kim。《皮层表面数据的持久图》。载于：医学影像信息处理国际会议。Springer。2009年，第386–397页。[EH10] Herbert Edelsbrunner 和 John Harer。《计算拓扑学：导论》。2010年1月。ISBN：978-0-8218-4925-5。DOI：10.1007/978-3-540-33259-6_7。[Ped+11] F. Pedregosa 等。《Scikit-learn：Python机器学习》。载于：Journal of Machine Learning Research 12（2011年），第2825–2830页。[She12] Donald R Sheehy。《Vietoris-Rips过滤的线性规模近似》。载于：第二十八届计算几何年度研讨会论文集。2012年，第239–248页。[BDG13] Jean-Daniel Boissonnat、Ramsay Dyer 和 Arijit Ghosh。《Delaunay三角剖分的稳定性》。载于：International Journal of Computational Geometry & Applications 23.04n05（2013年），第303–333页。[Kra+14] Miroslav Kram´ar 等。《量化颗粒系统中的力网络》。载于：Physica D: Nonlinear Phenomena 283（2014年），第37–55页。[Mar+14] Cl´ement Maria 等。《Gudhi 库：单纯复形与持久同调》。2014年6月。ISBN：978-3-662-44198-5。DOI：10.1007/978-3-662-44199-2_28。[TVA14] Andrew Tausz、Mikael Vejdemo-Johansson 和 Henry Adams。《JavaPlex：持久（上）同调研究软件包》。载于：ICMS 2014论文集。Han Hong 和 Chee Yap 编。Lecture Notes in Computer Science 8592。软件可于 http://appliedtopology. github.io/javaplex/ 获取。2014年，第129–136页。[Chi+15] Harish Chintakunta 等。《基于熵的持久条形码》。载于：Pattern Recognition 48.2（2015年），第391–401页。[BE16] Ulrich Bauer 和 Herbert Edelsbrunner。《Čech复形与Delaunay复形的莫尔斯理论》。载于：Transactions of the American Mathematical Society 369.5（2016年12月），第3741–3762页。ISSN：1088-6850。DOI：10.1090/tran/6991。URL：http://dx.doi.org/10.1090/tran/6991。[Qai+16] Talha Qaiser 等。《持久同调用于全切片组织学图像中的快速肿瘤分割》。载于：Procedia Computer Science 90（2016年）。第20届医学图像理解与分析会议（MIUA 2016），第119–124页。ISSN：1877-0509。DOI：10.1016/j.procs.2016.07.033.URL: https://www.sciencedirect.com/science/article/pii/S1877050916312133.[RSG16] Marco Tulio Ribeiro、Sameer Singh 和 Carlos Guestrin。《"为什么我应该信任你？"解释任何分类器的预测》。载于：第22届ACM SIGKDD知识发现与数据挖掘国际会议论文集。2016年，第1135–1144页。[VRT16] Vinay Venkataraman、Karthikeyan Natesan Ramamurthy 和 Pavan Turaga。《用于动作识别的吸引子持久同调》。载于：2016年 IEEE图像处理国际会议（ICIP）。2016年，第4150–4154页。DOI: 10.1109/ICIP.2016.7533141.[AAM17] Micha l Adamaszek、Henry Adams 和 Francis Motta。《随机循环动力系统》。载于：应用数学进展 83（2017年），第1–23页。ISSN: 0196-8858. DOI: 10.1016/j.aam.2016.08.007。URL: https://www.sciencedirect.com/science/article/pii/S0196885816300768.[Ada+17] Henry Adams 等人。《持久图像：持久同调的稳定向量表示》。载于：Journal of Machine Learning Research 18（2017年）。[AJ17] Aras Asaad 和 Sabah Jassim。《用于图像篡改检测的拓扑数据分析》。载于：Digital Forensics and Watermarking。Christian Kraetzer 等人编。Cham：Springer International Publishing，2017年，第 136–146 页。ISBN: 978-3-319-64185-0.

[Bau+17] Ulrich Bauer 等人。《Phat——持久同调算法工具箱》。载于：Journal of Symbolic Computation 78（2017年）。计算拓扑学算法与软件专辑，第 76–90 页。ISSN: 0747-7171. DOI: 10.1016/j.jsc.2016.03.008. URL: https://www.sciencedirect.com/science/article/pii/S0747717116300098.[KA17] Martin Krzywinski 和 Naomi Altman。《重要发现：分类与回归树》。载于：《自然·方法》14（2017年），第 757–758 页。[Ott+17] Nina Otter 等人。《持久同调计算路线图》。载于：EPJ 数据科学 6.17（2017年）。 DOI: 10.1140/epjds/s13688-017-0109-5.[MLW18] John Malik、Yu-Lun Lo 和 Hau-tieng Wu。《通过卷积神经网络量化心率变异性进行睡眠-觉醒分类》。载于：生理测量 39.8（2018年），第 085004 页。[TSB18] Christopher Tralie、Nathaniel Saul 和 Rann Bar-On。《Ripser.py：适用于 Python 的精简持久同调库》。载于：《开放源代码软件杂志》3.29（2018年9月），第 925 页。 DOI: 10.21105/joss.00925. URL: https://doi.org/10.21105/joss.00925.[ˇCuf20] Matija ˇCufar。《Ripserer.jl：灵活高效的 Julia 持久同调计算》。载于：Journal of Open Source Software 5.54（2020年），第 2614 页。 DOI: 10.21105/joss.02614. URL: https://doi.org/10.21105/joss.02614.[SW20] Mengbai Xiao、Simon Zhang 和 Hao Wang。GPU 加速计算 Vietoris-Rips 持久条形码。2020 年。电子预印本：arXiv:2003.07989.[ST20] Primoz Skraba 和 Katharine Turner。《持久图的 Wasserstein 稳定性》。载于：arXiv：《代数拓扑》（2020年）。[Vir+20] Pauli Virtanen 等人。《SciPy 1.0：用于 Python 科学计算的基础算法》。载于：《自然·方法》17（2020年），第 261–272 页。 DOI: 10.1038/s41592-019-0686-2.[Bau21] Ulrich Bauer。《Ripser：高效计算 Vietoris–Rips 持久条形码》。载于：Journal of Applied and Computational Topology 5.3（2021年），第 391–423 页。[Chu+21] Yu-Min Chung 等人。《持久同调方法在心率变异性分析中的应用及其在睡眠-觉醒分类中的应用》。载于：Frontiers in Physiology 12（2021年）。 ISSN: 1664- 042X. DOI: 10.3389/fphys.2021.637684. URL: https://www.frontiersin.org/article/10. 3389/fphys.2021.637684.[Rou23] Vincent Rouvreau。《Alpha 复形》。载于：GUDHI 用户与参考手册。3.7.1。GUDHI 编辑委员会，2023年。 URL: https://gudhi.inria.fr/doc/3.7.1/group__alpha__complex. html.[Misa] Amish Mishra。将合成数据分类为形状类别。 URL: https://github.com/ amish-mishra/TDA_shape_classification_using_DR.[Misb] Amish Mishra。Delaunay-Rips。 URL: https://github.com/amish-mishra/cechmate-DR.[Misc] Amish Mishra。ML-Del-Rips-sleep-wake-classification。 URL: https://github.com/amish-mishra/ ML-Del-Rips-sleep-wake-classification.[Mor] Dmitriy Morozov。Dionysus。 URL: https://pypi.org/project/dionysus/.[Nan] Vidit Nanda。Perseus，持久同调软件。访问日期：2023年3月1日。 URL: http:// www.sas.upenn.edu/~vnanda/perseus.[Rei] Jan Reininghaus。DIPHA（一种分布式持久同调算法）。 URL: https://github. com/DIPHA/dipha.[Sau+] Nathaniel Saul 等人。Persim。许可证：MIT。 URL: https://persim.scikit-tda.org/en/latest/.[TS] Chris Tralie 和 Nathaniel Saul。Cechmate。许可证：MIT。 URL: https://github.com/scikit-tda

/

Cechmate。